



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Path Intergal Evidence

Citation for published version:

Kitching, T & Taylor, A 2015, 'Path Intergal Evidence', *Monthly Notices of the Royal Astronomical Society*, vol. 452, pp. 1519.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Monthly Notices of the Royal Astronomical Society

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Path-integral evidence

T. D. Kitching^{1★} and A. N. Taylor²

¹*Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK*

²*SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK*

Accepted 2015 June 3. Received 2015 June 3; in original form 2014 December 10

ABSTRACT

Here we present a Bayesian formalism for the goodness of fit that is the evidence for a fixed functional form over the evidence for all functions that are a general perturbation about this form. This is done under the assumption that the statistical properties of the data can be modelled by a multivariate Gaussian distribution. We use this to show how one can optimize an experiment to find evidence for a fixed function over perturbations about this function. We apply this formalism to an illustrative problem of measuring perturbations in the dark energy equation of state about a cosmological constant.

Key words: cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Finding an objective Bayesian measure for a goodness of fit of a function to some data is problematic because when one calculates a Bayes factor (the ratio of pieces of evidence for particular models) the assumption of at least two models must be made. To circumvent this assumption one can ask what the evidence for a particular (singular) model is over all possible perturbations about that model. We present such an evidence ratio here, under some assumptions, which is calculated using a path-integral methodology (an extension of the formalism described in Taylor & Kitching 2010; Kitching & Taylor 2011) that marginalizes over all functional perturbations about a fixed function. This then provides goodness of fit that is the evidence for the fixed function over all other functions. To achieve this we assume that the statistical properties of the data can be modelled by a multivariate Gaussian distribution. Using this ‘path-integral evidence’ we also define a utility function that includes a path-integral over the set of perturbed functions and future data, this can be used to calculate the expected value of evidence for a fixed function over perturbations about that function for a future experiment.

We apply this methodology to the case of determining if cosmological experiments can determine the true dark energy equation of state, as a function of redshift, $w(z)$, from perturbations about this function. We use as a simple example an experiment that aims to measure $w(z)$ from changes in the Hubble parameter. This is a particularly important question because dark energy accounts for the majority of the mass-energy content of the Universe and its nature is entirely unknown; explanations include proposals to modify general relativity on cosmological scales and the addition of new fundamental fields, amongst others. The strongest prediction for $w(z)$ is that it has a value of $w(z) = -1$ for all redshifts, which would indi-

cate that dark energy is a vacuum energy, or additional gravitational constant known as the ‘cosmological constant’. Measuring any deviation from this functional form, at any redshift, would necessitate the need for new physics beyond the standard models of cosmology and/or particle physics. This article is arranged as follows: in Section 2 we present the general methodology, in Section 3 we present an application to dark energy experiments and in Section 4 we present conclusions.

2 METHOD

We begin with a function that depends on a variable x and some parameters θ , that has a fixed (fiducial) functional form $f_F(x|\theta_0)$, where the parameters take particular values θ_0 . We perturb this such that any function could be written as a perturbation away from the fiducial

$$f(x) = f_F(x|\theta_0) + \delta_f(x). \quad (1)$$

These functions exist in a set M that contains the fiducial function, and all perturbations about that function. We can write a log-likelihood for the parameters θ in the case that the data can be modelled by a Gaussian distribution

$$-2\mathcal{L} = \Delta D^T \mathbf{C}_0^{-1} \Delta D + \ln |\mathbf{C}_0|, \quad (2)$$

where $\Delta D = \mu(\theta) - D$ is the difference of the mean, that is a function of the parameters θ , and the data D and we define a covariance \mathbf{C}_0 . We denote the Fisher matrix for the free parameters θ as $F_{\theta\theta}$ calculated using the covariance \mathbf{C}_0 .

From Kitching & Taylor (2011) we can write a likelihood for the parameters θ accounting for a path-integral marginalization over the functional behaviour of $\delta_f(x)$, by changing the covariance such that

$$\mathbf{C}_M = [\mathbf{C}_0^{-1} - \mathbf{C}_0^{-1} \mathcal{F} \mathbf{C}_0^{-1}]^{-1}, \quad (3)$$

* E-mail: t.kitching@ucl.ac.uk

where

$$\mathcal{F} = \int dx dx' \frac{\delta \mu}{\delta \delta_f(x)} F_{\delta_f \delta_f}^{-1}(x, x') \frac{\delta \mu}{\delta \delta_f(x')} \quad (4)$$

the mean is now a function of θ and $\delta_f(x)$, and $F_{\delta_f \delta_f}(x, x')$ are the elements of the functional Fisher matrix for the perturbations δ_f marginalized over the parameters θ . This is the ‘self-calibration’ case discussed in Kitching & Taylor (2011) where the prior in function-space is a flat top hat; for a detailed discussion of flat functional priors see Bornkamp (2011).

In the case that a functional prior is included over the perturbed function-space $\delta_f(x)$ the covariance reduces (via the Woodbury matrix identity) to

$$\mathbf{C}_M = \mathbf{C}_0 + \mathcal{G}, \quad (5)$$

where

$$\mathcal{G} = \int dx dx' \frac{\delta \mu}{\delta \delta_f(x)} C_{P, \delta_f \delta_f}(x, x') \frac{\delta \mu}{\delta \delta_f(x')}, \quad (6)$$

where C_P is the covariance of a Gaussian functional prior. Either the flat or Gaussian-prior cases (equations 3 and 5) can be used to construct a new Fisher matrix for the parameters θ , now taking into account the marginalization over perturbations δ_f about the fixed function. As discussed in Taylor & Kitching (2010) the flat prior case is different to a Gaussian prior with large variance, they have different functional weighting that means the latter allows for the use of the Woodbury identity whereas the former does not.

We can now extend the Bayesian evidence formalism of Taylor & Kitching (2010), to include the marginalization over the perturbed functional behaviour. Referring to Taylor & Kitching (2010, equation 51) we write the evidence, assuming that the data can be modelled using Gaussian distributions, including marginalization over all perturbations about the fiducial function

$$\ln E^M = \Delta D^T \mathbf{C}_M^{-1} \Delta D + \ln |\mathbf{C}_M| + 2 \ln(V^M \sqrt{|F^M|}) - N_M \ln 2\pi, \quad (7)$$

where \mathbf{C}_M is defined either in equation (5) or (3) depending on the prior. V^M is the volume over which the prior distribution is defined. F^M is the Fisher matrix for the fixed parameters after marginalization over the perturbed case, this is defined as the Schur complement of the full Fisher matrix F i.e. $F_{ij}^M = F_{ij} - F_{i\alpha} F_{\alpha\beta}^{-1} F_{\beta j}$, where α and β here represent function elements of the total Fisher matrix, and Roman letters represent parameters of the fixed function. We can also write the evidence for the fixed function by referring to Taylor & Kitching (2010, equation 54)

$$\ln E^0 = \Delta D^T \mathbf{C}_0^{-1} \Delta D + \ln |\mathbf{C}_0| + 2 \ln(V^0 \sqrt{|F^0|}) - N_0 \ln 2\pi, \quad (8)$$

where \mathbf{C}_0 is the covariance not marginalizing over the perturbations $\delta_f(x)$, and F^0 is the part of the Fisher matrix for the parameters of the fixed function only. V^0 is the volume over which the prior distribution is defined, in the fixed function case. $N_0 = N_D - N_P$ is the number of degrees of freedom where N_D is the number of data points and N_P is the number of free parameters. In the case of the perturbed case $N_M \rightarrow -\infty$. We address the infinity below.

By combining the expressions for the evidence in the fixed plus perturbed ($F + P$) case $f(x) = f_F(x) + \delta_f(x)$, and the fixed case $f(x) = f_F(x)$ only (F), we can write a Bayes factor that quantifies the

ability of the data to distinguish these models

$$\begin{aligned} \mathcal{B}_{F/(F+P)} &= \ln E^0 - \ln E^M \\ &= \Delta D^T (\mathbf{C}_0^{-1} - \mathbf{C}_M^{-1}) \Delta D \\ &\quad + \ln |\mathbf{C}_0 \mathbf{C}_M^{-1}|. \end{aligned} \quad (9)$$

This is the evidence for the fixed model over the fixed plus the perturbed case.

The final terms in the evidence, the Occam factor, do not appear in this evidence ratio. This is because, in the perturbed case the parameter space volume tends to infinity. There are two motivations that could be used to justify the removal of these terms. First the removal of the Occam term can be considered as a practical measure. The Bayes factor would then not account for the Occam factor, and is therefore not a true Bayesian measure, but does still capture the change to the posterior caused by the increase in covariance as a result of marginalization over the larger parameter (function) space.

Alternatively it can be justified by modifying the prior volume V^M in the perturbed case such that in the two different models this term cancels. This is achieved by scaling the volume V^M (a hypersphere with radii scaled using the ratio of the expected Fisher matrices), over which we allow the prior in the perturbed case to be defined, by the prior volume of the unperturbed case

$$V^M = V^0 (2\pi)^{(N_M - N_0)/2} \left[\frac{F^0}{F^M} \right]^{1/2}. \quad (10)$$

The above expression is the volume of the hypersphere V^M scaled from V^0 . Note that the integral over prior (outside this volume) can tend to infinity but that in this case, because we always scale with respect to V^0 in taking the evidence ratio, the Occam factor terms cancel to zero. This is equivalent to setting a prior on the parameter volume, as a result this metric will not be sensitive to penalties induced from an unexpected increase in volume. Without some justification for the removal of the Occam factor the evidence ratio would result in an infinite result (the marginalized case having an infinite number of degrees of freedom), but we note that the scaling is convergent, and so asymptotically should behave correctly.

2.1 Expectation value

We now explore the expectation value of the above statistic, and show how we can forecast this Bayes factor for experimental optimization.

For forecasting purposes we need to integrate over the future data vector, as described in Kitching & Amara (2009) and Trotta, Kunz & Liddle (2011) where an ‘expected utility’ or ‘figure of merit’ is defined as an integral over parameters (or functional integrals in our case), and data vectors. The utility function here case is the log Bayes factor for the fixed model over the perturbed case $\mathcal{B}_{F/(F+P)}$, given in equation (9), where we have already performed the integrals over functions and parameters. The final integral over the data vector arrives at an expression for the expected evidence ratio

$$\langle \mathcal{B}_{F/(F+P)} \rangle = 1 - \text{Tr} (\mathbf{C}_0 \mathbf{C}_M^{-1}) + \ln |\mathbf{C}_0 \mathbf{C}_M^{-1}|. \quad (11)$$

We recognize the two terms above as: (i) the fractional change in the covariance, (ii) the fractional change in the volume of the covariance due to the functional marginalization. This expression now quantifies the expected ability of an experiment to distinguish a fixed function $f_F(x)$ from the fixed plus perturbed case $f_F(x) + \delta_f(x)$.

Interestingly, this in fact has a similar form to the *Itakura–Saito distance* that is a measure of the ‘perceptual difference between an original spectrum P and an approximation \hat{P} of that spectrum’, of use in signal processing, but here we derive this from first principles. More generally equation (11) is similar to the Kullback–Leibler divergence $D_{\text{KL}}(p_1, p_2) = \int dx p_1(x) \ln(p_2(x)/p_1(x))$, which is a measure of the difference between two distributions, in our case the probability of the true function and the probability marginalized over all functions. The Kullback–Leibler divergence was investigated as a utility function or ‘figure of merit’ for cosmology in Amara & Refregier (2014). However this only looked at simple functional changes, while we generalize to the continuous functional case and derive from first principles.

We can also calculate the evidence for the fixed model over the perturbed case

$$\langle \mathcal{B}_{F/P} \rangle = -\ln [e^{-(\mathcal{B}_{F/(F+P)})} - 1]. \quad (12)$$

Here we assume that the pieces of evidence for the fixed and perturbed models are additive, $p(D|F + P) = p(D|F) + p(D|P)$, where the probability $p(D|F + P)$ corresponds to the ‘probability of the fixed OR perturbed’ case (i.e. any function). This is true because F and P are mutually exclusive; here we define the perturbed model-space as all functions *except* the fiducial case (another way of formulating this is that fiducial function F is a special case within the space of all functions, in that its prior volume is zero). This is an expression for the expectation value of the evidence for the true function over perturbations away from that true function. To calculate this we make the approximation that $\langle \mathcal{B}_{F/(F+P)} \rangle = \ln[(E^0/(E^0 + E^M))]$.

This statistic can be used to optimize experiments so that they will have the ability to distinguish a fixed function from perturbations about that function. In practice one would fit a free functional form to the data (see e.g. Crittenden et al. 2012). For calculation purposes one must assume a ‘fiducial’ model for the fixed function $f_F(x)$ around which to calculate the Fisher matrices and covariances. However, we note that under the assumption that the data can be modelled by a multivariate Gaussian the expressions above are *independent* of the choice of fiducial function (see Taylor & Kitching 2010).

3 EXAMPLE APPLICATION

We now apply this methodology to the dark energy equation of state $w(z)$, where we have a model that includes a fixed function of $w_F(z)$ and we include perturbations about this model

$$w(z) = w_F(z) + \delta_w(z), \quad (13)$$

where the function is over redshift z .

We look at an illustrative example in this paper as a proof of concept, where we consider a measure of the Hubble parameter $H(z)$ which is related to $w(z)$ via the following equation

$$H(z) = H_0 \left[\Omega_M(1+z)^3 + \Omega_X(1+z)^3 \int_0^z dz' [1+w(z')]/(1+z') \right]^{1/2}, \quad (14)$$

where H_0 is the current rate of expansion, Ω_M and Ω_X are the dimensionless matter and dark energy densities respectively and $w(z)$ is the dark energy equation of state. We consider a hypothetical experiment that measures $H(z)$ in ten redshift bins between redshifts $0 \leq z \leq 2$ with an accuracy of σ_H in each redshift bin. We use a fiducial cosmology that assumes $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, $\Omega_X = 0.7$ and $w_F(z) = -1 \forall z$. We assume a functional prior on $w(z)$ that is diagonal in redshift $C_{P\delta_{w(z)}\delta_{w(z')}} = \sigma_P^2 \delta_{zz'}^K$. In Fig. 1 we show an example where the measurement on $H(z)$ has an error bar of $\sigma_H = 10^{-3}$, all errors have the same units as H .

It is worth spending some time examining the generic behaviour seen in Fig. 1. The theory prior is a functional prior around the fixed function. For the Bayes factor $\langle \mathcal{B}_{F/(F+P)} \rangle$ what we find is that as the theory prior approaches infinity the Bayes factor tends to minus infinity, this means that data would favour a perturbed model. As the theory prior tends to zero the Bayes factor $\langle \mathcal{B}_{F/(F+P)} \rangle \rightarrow 1$, as expected from equation (11), this means that if one performed this experiment with a small theoretical prior there would be no change in evidence $\langle \mathcal{B}_{F/(F+P)} \rangle$ i.e. with this prior *it would not be worth doing this experiment* because the space of models tested is already well constrained by the prior, and the experiment is expected to favour the fixed function.

The Bayes factor $\langle \mathcal{B}_{F/P} \rangle$ has an interesting and important behaviour. When $\langle \mathcal{B}_{F/(F+P)} \rangle = 0$ there is a singularity where the expected evidence for the fixed model tends to infinity. This boundary is of particular importance. If one performs an experiment with a

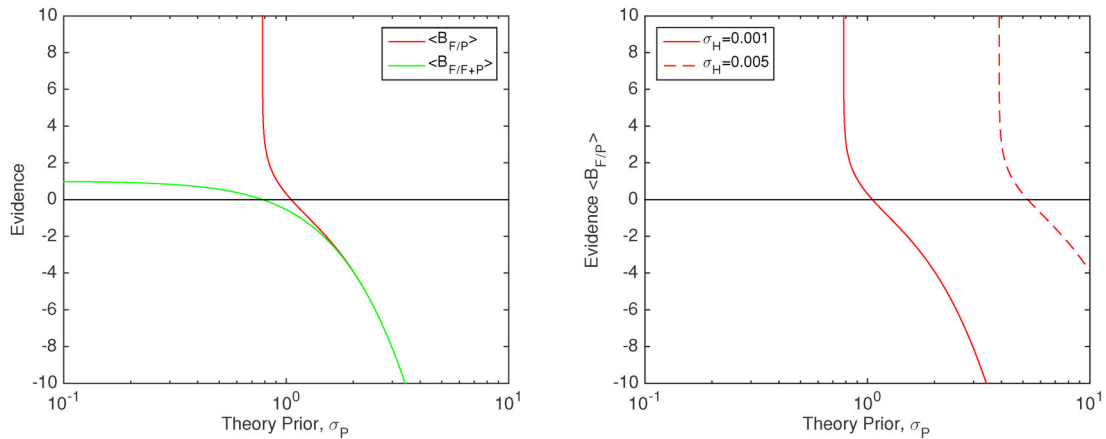


Figure 1. Left: the expected evidence $\langle \mathcal{B}_{F/P} \rangle$ (fixed versus perturbed) and $\langle \mathcal{B}_{F/(F+P)} \rangle$ (fixed versus all) as a function of the pre-experimental theoretical prior. This is for the illustrative example of measuring a perturbation about a fixed value of the dark energy equation of state $w(z) = -1$ using measurements of the Hubble parameter. Shown are expected evidence for an expected error on the measurements of $\sigma_H = 10^{-3}$. Right: the expected evidence $\langle \mathcal{B}_{F/P} \rangle$ (fixed versus perturbed) for two different values of the expected error on H .

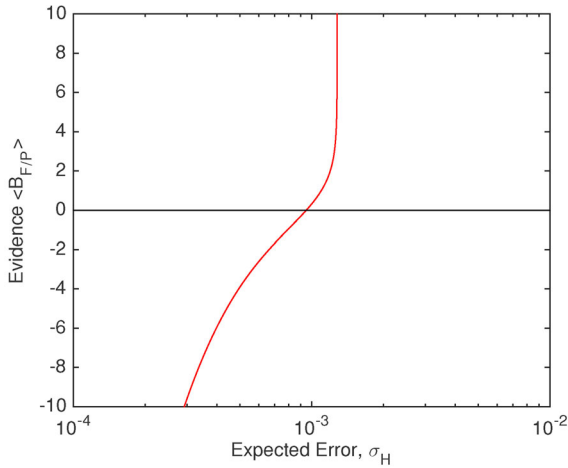


Figure 2. The expected evidence $\langle \mathcal{B}_{F/P} \rangle$ (fixed versus perturbed) as a function of the expected error bar on the Hubble parameter σ_H ; here we fix the pre-experimental theory prior to $\sigma_P = 1$.

pre-experimental theoretical prior smaller than a particular value (leftward of this singularity in Fig. 1) then the experiment is likely to return no change in evidence over that prior, and is likely to favour the fixed function. However, if one performs an experiment with a pre-experimental prior that is larger than the experiment is likely to return a change in evidence (either a weaker preference for the fixed function, or for the perturbed case). In colloquial terms this singularity demarks, for a given expected experimental error, and a given pre-experimental theoretical prior, whether the experiment is ‘worth doing’.

To explore this concept further we show in Fig. 1 the case of two different expected error bars on $H(z)$. Consider in this figure the case that before doing the experiment the theoretical prior about $w(z)$ was $\sigma_P = 1$, in this case it would be worth doing an experiment with an expected error of $\sigma_H = 10^{-3}$ but it would not be worth doing an experiment with $\sigma_H = 5 \times 10^{-3}$.

Furthermore, as the error bar decreases the expected evidence for of finding a perturbation increases ($\langle \mathcal{B}_{F/P} \rangle$ becomes more negative). If one is rightwards of the singularity in Fig. 1 then an experiment may be worth doing, but it can be highly likely that it will find no evidence for a perturbation. As the experimental error bar decreases one becomes more likely to find evidence for a perturbation over the fixed case (negative values of $\langle \mathcal{B}_{F/P} \rangle$). There is a point where the evidence of the fixed function becomes zero $\langle \mathcal{B}_{F/P} \rangle = 0$, this occurs when $\langle \mathcal{B}_{F/(F+P)} \rangle = -\ln(2)$ i.e. that there is equal evidence for the fixed and perturbed case ($\langle \mathcal{B}_{F/(F+P)} \rangle = \ln[E_F/(E_F + E_P)] = \ln(1/2)$).

We explore this in Fig. 2, where the expected evidence $\langle \mathcal{B}_{F/P} \rangle$ is shown as a function of the expected experimental error σ_H . When the expected error is larger there is no expected change in evidence due to the experiment. As the expected error decreases there is a regime where the experiment becomes worth doing – but the

result is expected to confirm the non-perturbed model (i.e. given the theoretical prior the data is unlikely to provide evidence against the non-perturbed model). However, when the expected error is small enough, in this case $\sigma_H \lesssim 10^{-3}$, the experiment is worth doing and is expected to return evidence for a perturbed model over the non-perturbed case – if such a perturbation exists.

4 CONCLUSION

Here we present, under the assumption that the data can be modelled by a multivariate Gaussian distribution, a Bayesian goodness of fit which compares the evidence for a fixed functional form to the total evidence for all functions that are not the fixed functional form. We then define a utility function using this metric that marginalizes over all functional perturbations about a fixed model.

We apply this using an illustrative example of determining the dark energy equation state through measurements of the Hubble parameter. We find that the expected evidence provides a clear metric, given a pre-experimental theoretical prior and expected error bar, that can quantify if an experiment is likely to produce a change in the evidence; answering the question is an experiment ‘worth doing’? We find that as the expected error bar for a planned experiment decreases the expected evidence for finding a perturbation can become significant, and we suggest that this metric may be useful in experimental design. For our illustrative example we show that in order for an experiment to expect to find strong evidence for a perturbation about $w(z) = -1$ a measurement of the Hubble parameter only would need to have errors of $\sigma_H \lesssim 10^{-3}$ over 10 redshift bins between $0 \leq z \leq 2$.

This formalism should be applicable to any case of experimental design that seeks to maximize the expectation of finding perturbations about a functions, and as such can be used as a model-independent tool for experimental optimization.

ACKNOWLEDGEMENTS

TDK is supported by a Royal Society University Research Fellowship. We thank an anonymous referee for comments that improved an earlier version of this paper.

REFERENCES

- Amara A., Refregier A., 2014, Phys. Rev. D, 89, 083501
- Bornkamp B., 2011, preprint ([arXiv:1110.4400](https://arxiv.org/abs/1110.4400))
- Crittenden R. G., Zhao G.-B., Pogosian L., Samushia L., Zhang X., 2012, J. Cosmol. Astropart. Phys., 048, 2012
- Kitching T. D., Amara A., 2009, MNRAS, 398, 2134
- Kitching T. D., Taylor A. N., 2011, MNRAS, 410, 1677
- Taylor A. N., Kitching T. D., 2010, MNRAS, 408, 865
- Trotta R., Kunz M., Liddle A. R., 2011, MNRAS, 414, 2337

This paper has been typeset from a \LaTeX file prepared by the author.